PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-090970
(43)Date f publication of application : 04.04.1997

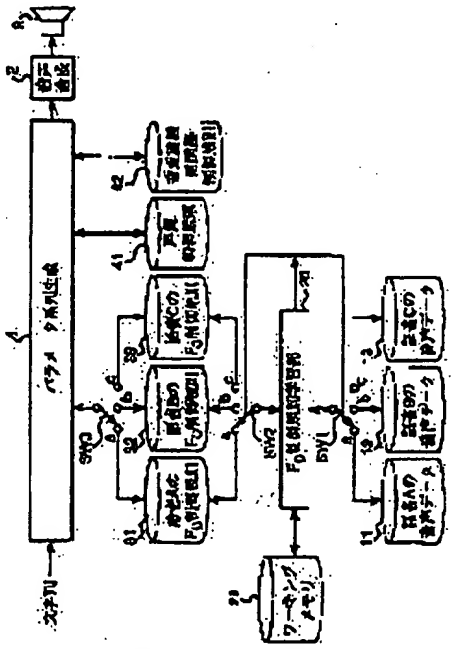
(51)Int.Cl. G10L 3/00
G10L 5/02

(21)Application number : 07-241460
(22)Date of filing : 20.09.1995

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : HIRAI TOSHIO
KOSAKA YOSHINORI
HIGUCHI NORIO

(54) SPEECH SYNTHESIS DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To synthesize speech of one designated speaker by c ntrolling a size of an accent phrase based on an accent type of an a cent phrase of a speech synthesis object and the position of an accent phrase in sentences.
SOLUTION: A parameter system generation part 1 by which a feature parameter system for a speech synthesis is generated based on the input character string, and a speech synthesis part 2 outputting to speaker 3 which generates a speech signal based on the generated feature parameter system are provided. And an inputted character string is converted into the voice of a prescribed speaker by using F0 control rules 31-33 controlling the pitch frequency of the voice made for each speaker. The F0 Control rules 31-33 are rules which control the size of the phrase based on the number of the mora of the phrases of the speech synthesis object and the numbers of the morae of the phrase preceding to this phrase, and which control the size of the accent phrase and the pitch frequency of the speech based on the accent type of the accent phrase of the voice synthesis object and the position of the accent phrase in the sentence.



EGAL STATUS

[Date f request for examination] 20.09.1995
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date f final disposal for application]
[Patent number] 2880433
[Date of registration] 29.01.1999
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号
特開平9－90970
(43)公開日　平成９年(1997)４月４日

(51)Int.Cl.⁶		識別記号	庁内整理番号	FI			技術表示箇所
G10L	3/00			G10L	3/00	H	
	5/02				5/02	J	

審査請求 有　請求項の数３ ＯＬ （全 11 頁）

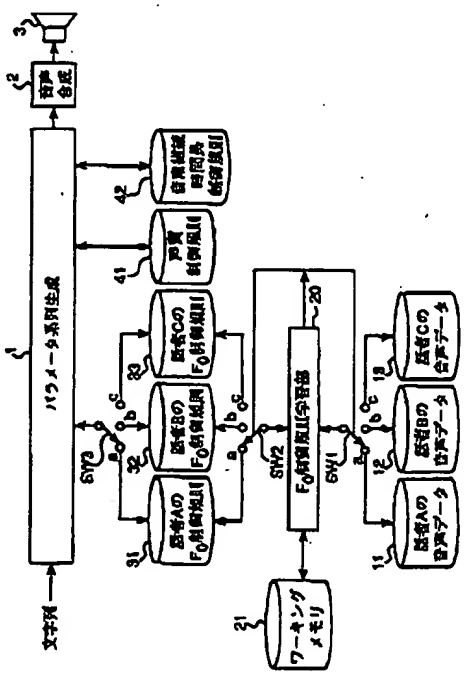
(21)出願番号　特願平7－241460
(22)出願日　平成７年(1995)９月20日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地
(72)発明者　平井　俊男
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内
(74)代理人　弁理士　青山　葆　（外２名）

最終頁に続く

(54)【発明の名称】　音声合成装置

(57)【要約】
【課題】　ある指定された１人の話者の音声を合成することができる音声合成装置を提供する。
【解決手段】　話者毎に作成された音声のピッチ周波数を制御する制御規則を用いて入力された文字列を予め指定された話者の音声に変換する。制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する規則であり、音声データに基づいて音声のピッチ周波数のパターンを抽出し、抽出された音声のピッチ周波数のパターンに基づいて臨界制御モデルによる分析法を用いて臨界制御モデルのモデルパラメータを発生し、音声のピッチ周波数を制御する制御規則を生成する。



1

【特許請求の範囲】

【請求項１】　入力された文字列に基づいて音声を合成する音声合成装置において、

話者毎に作成された音声のピッチ周波数を制御する制御規則を用いて入力された文字列を予め指定された話者の音声に変換する変換手段を備えたことを特徴とする音声合成装置。

【請求項２】　上記制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する規則であることを特徴とする請求項１記載の音声合成装置。

【請求項３】　上記音声合成装置はさらに、上記制御規則を生成する学習手段を備え、

上記学習手段は、

音声データに基づいて音声のピッチ周波数のパターンを抽出する抽出手段と、

上記抽出手段によって抽出された音声のピッチ周波数のパターンに基づいて臨界制御モデルによる分析法を用いて上記臨界制御モデルのモデルパラメータを発生する発生手段と、

上記抽出手段によって抽出された音声のピッチ周波数のパターンと、上記発生手段によって発生された上記臨界制御モデルのモデルパラメータとに基づいて、音声のピッチ周波数を制御する制御規則を生成する生成手段とを備えたことを特徴とする請求項１又は２記載の音声合成装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、入力された文字列に基づいて音声を合成する音声合成装置に関する。

【０００２】

【従来の技術】音声の基本周波数であるピッチ周波数（以下、Ｆ０周波数という。）のモデル化には、従来から少ないパラメータ数で効率良くＦ０周波数の時系列のパターン（以下、Ｆ０パターンという。）をパラメータ化することが可能な重畳型モデルが用いられている（例えば、従来文献１「藤崎ほか，"日本語単語アクセントの基本周波数パタンとその生成機構のモデル"，日本音響学会論文誌，Ｖｏｌ．２７．Ｎｏ．９，ｐｐ．４４５－４５３，１９７１年９月」参照。）この重畳型モデルでは、Ｆ０パターンを句頭から句末にかけて緩やかに下降するフレーズ成分（話調成分とも呼ばれる。）とアクセント句に対応するアクセント成分の和として捉える。重畳型モデルによるＦ０パターンのパラメータ化には、次のような利点がある。

（１）モデルで用いられる自由パラメータ数が少なく、

2

統計分析によるＦ０制御の最適化が容易である。

（２）Ｆ０パターンをフレーズ成分とアクセント成分の２つの成分に分離するので、最適化するので、最適化の結果得られる制御規則の解釈が比較的容易である。

【０００３】また、規則合成音声の多様化を図るための、普通調、コマーシャル調、朗読調の３つの発話様式の間の変換規則（以下、従来例という。）が、例えば従来文献２「阿部ほか，"発話様式の変化とその評価"，日本音響学会講演論文集，３－Ｐ－１８，１９９３年１０月」において提案されている。この従来例では、フォルマント周波数と継続時間と基本周波数及びパワーのパラメータを変換することにより、普通調、コマーシャル調、朗読調の各音声と、普通調からコマーシャル調へ変換した音声と、普通調から朗読調へ変換した音声の計５つの発話様式を準備して、それらの類似性について評価している。

【０００４】

【発明が解決しようとする課題】しかしながら、上述の従来例では、発話様式の変換を対象としているが、話者性を考慮せずに音声合成している。すなわち、アクセント型が異なるとアクセントの高さが個人により異なり、従来例では、ある指定された１人の話者の音声を合成することはできない。

【０００５】本発明の目的は以上の問題点を解決し、ある指定された１人の話者の音声を合成することができる音声合成装置を提供することにある。

【０００６】

【課題を解決するための手段】本発明に係る請求項１記載の音声合成装置は、入力された文字列に基づいて音声を合成する音声合成装置において、話者毎に作成された音声のピッチ周波数を制御する制御規則を用いて入力された文字列を予め指定された話者の音声に変換する変換手段を備えたことを特徴とする。

【０００７】また、請求項２記載の音声合成装置は、請求項１記載の音声合成装置において、上記制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する規則であることを特徴とする。

【０００８】さらに、請求項３記載の音声合成装置は、請求項１又は２記載の音声合成装置において、さらに、上記制御規則を生成する学習手段を備え、上記学習手段は、音声データに基づいて音声のピッチ周波数のパターンを抽出する抽出手段と、上記抽出手段によって抽出された音声のピッチ周波数のパターンに基づいて臨界制御モデルによる分析法を用いて上記臨界制御モデルのモデルパラメータを発生する発生手段と、上記抽出手段によ

って抽出された音声のピッチ周波数のパターンと、上記発生手段によって発生された上記臨界制御モデルのモデルパラメータとに基づいて、音声のピッチ周波数を制御する制御規則を生成する生成手段とを備えたことを特徴とする。

【０００９】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。図１は、本実施形態のＦ０周波数を制御するＦ０制御規則を生成するＦ０制御規則学習部２０を備えた音声合成装置のブロック図である。図１において、本実施形態の音声合成装置は、入力される文字列に基づいて、逐択的に接続される１人の話者のＦ０制御規則（３１，３２，３３のうちの１つ）と、声質制御規則４１と、音素継続時間長制御規則４２とを用いて音声合成のための特徴パラメータ系列を生成するパラメータ系列生成部１と、生成された特徴パラメータ系列に基づいて音声信号を発生してスピーカ３に出力する音声合成部２とを備える。本実施形態においては、特に、話者毎に作成された音声のピッチ周波数を制御するＦ０制御規則３１，３２，３３を用いて入力された文字列を予め指定された話者の音声に変換することを特徴とし、上記Ｆ０制御規則３１，３２，３３は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する規則である。

【００１０】Ｆ０制御規則学習部２０には、詳細後述するＦ０制御規則学習処理を実行するときのワークエリアとして用いるワーキングメモリ２１が接続される。また、Ｆ０制御規則学習部２０には、スイッチＳＷ１を介して、話者Ａ，Ｂ，Ｃの音声データ１１，１２，１３のうちの１つが逐択的に接続される一方、スイッチＳＷ２を介して、話者Ａ，Ｂ，ＣのＦ０制御規則３１，３２，３３のうちの１つが逐択的に接続される。これらのスイッチＳＷ１，ＳＷ２の切り換えはＦ０制御規則学習部２０によって、同一の話者の音声データとＦ０制御規則が同時に接続されるように連動して制御される。さらに、パラメータ系列生成部１には、スイッチＳＷ３を介して、話者Ａ，Ｂ，ＣのＦ０制御規則３１，３２，３３のうちの１つが逐択的に接続される。このスイッチＳＷ３の切り換えは、操作者により音声合成した話者のＦ０制御規則を逐択するように行われる。また、パラメータ系列生成部１には、詳細後述する従来の声質変換制御規則４１と従来の音素継続時間長制御規則４２とが接続される。

【００１１】本実施形態において、音声データ１１，１２，１３と、ワーキングメモリ２１と、Ｆ０制御規則３１，３２，３３と、声質制御規則４１と、音素継続時間長制御規則４２とは、例えば、ハードディスクなどのメモリで構成される。また、Ｆ０制御規則学習部２０と、パラメータ系列生成部１とは、例えばデジタル電子計算機で構成される。

【００１２】図２は、図１のＦ０制御規則学習部２０によって実行されるＦ０制御規則学習処理を示すフローチャートである。まず、ステップＳ１では、音声データ１１，１２，１３内の音声データに基づいてＦ０パターンを抽出した後、ステップＳ２において、抽出されたＦ０パターンに基づいて臨界制御モデルによる分析法を用いて上記臨界制御モデルのモデルパラメータを発生する。さらに、ステップＳ３で、抽出されたＦ０パターンと、臨界制御モデルのモデルパラメータとに基づいて、所定の制御要因に注目して、音声のピッチ周波数を制御する制御規則を生成する。ここで、制御要因とは、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数と、音声合成対象のアクセント句のアクセント型と、上記アクセント句の文章内の位置であり、Ｆ０制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する。次いで、上記各ステップの処理の詳細について説明する。

【００１３】まず、ステップＳ１の処理について述べる。音声データ１１，１２，１３にはそれぞれ、１人の話者の読み上げ文（発声音声文ともいう。）の音声信号のデータを含む。このステップＳ１では、この音声信号のデータに対して、Ａ／Ｄ変換とＬＰＣ分析を行って特徴パラメータデータを抽出した後、抽出した特徴パラメータデータに基づいて、例えば公知の臨界制動モデルによる分析法（例えば、従来文献３「藤崎ほか，“Analysis of voice fundamental frequency contours for declarative sentences of Japanese （日本語平叙文の基本周波数パターンの分析）”，日本音響学会論文誌，（Ｅ），Ｖｏｌ．５，Ｎｏ．４，ｐｐ．２３３－２４４，１９８４年４月」参照。）により分析しかつＦ０パターンとモデルパラメータとを検出して、音素単位、アクセント句単位及びフレーズ単位でラベリングすることより生成する。ここで、特徴パラメータデータは、対数パワー、１６次ケプストラム係数、Δ対数パワー、及び１６次Δケプストラム係数を含み、モデルパラメータとは、アクセント指令と、フレーズ指令とを含み、この中で、アクセント句境界の情報を含む。上記分析では、Ｆ０周波数の緩やかな下降成分であるフレーズ成分とＦ０周波数の局所的な起伏を示すアクセント成分に分解される。上記臨界制動モデルでは、フレーズ成分、アクセント成分はそれぞれフレーズ指令、アクセント指令に対す

る臨界制動２次線形系の応答として捉える。各指令の精密なタイミングと大きさは、音素ラベリング情報、アクセント句情報、フレーズ境界情報から得られるフレーズ指令、アクセント指令のおおよそのタイミングをもとに自動的な合成による解析（Ａｎａｌｙｓｉｓ－ｂｙ－Ｓｙｎｔｈｅｓｉｓ）を用いて求めることができる。

【００１４】次いで、ステップＳ２の処理について述べる。上述の重畳型制御モデルの１つとして、藤崎により研究提案されてきた藤崎モデルが知られている（例えば、従来文献１参照。）。この藤崎モデルを用いたパラメータ化には、従来、山登り法が用いられてきた（例えば、従来文献３参照。）。すなわち、藤崎モデルのすべての自由パラメータを変化させ、Ｆ₀パターンの平均推定２乗誤差を最小にするパラメータの組を、そのＦ₀パターンの分析結果とするものである。これは、パラメータの総数を探索空間次元数とする探索問題ととらえることができる。従来は、フレーズ指令に関しては角周波数、入力時点、及び大きさ、アクセント指令に関しては角周波数、立ち上がり時点、立ち下がり時点、及び大きさを自由パラメータとして取り扱っていたため，探索空間は（３Ｉ＋４Ｊ）次元（ここで、Ｉはフレーズ指令の数であり、Ｊはアクセント指令の数である。）であった。これらのパラメータのうち、アクセント指令の大きさは、Ｆ₀周波数の実測値と他のパラメータを与えれば、最小２乗法を用いて一意に求めることが可能でき、探索空間の次元数をＪだけ下げることにより計算時間を短縮することができる。この方法では、各時点でのＦ₀周波数の値の信頼性（ここでは、音声からＦ₀周波数を計算する際に得られる自己相関関数の極大値）を各時点でのＦ₀周波数の推定誤差評価の重み付けに用いることができるよう定式化している。これは、音声データから得られるＦ₀周波数の値の信頼性が各時点で一様ではないことに対応するためのものである。本実施形態においては、この方法を用いた山登り法によりＦ₀パターンのパラメータ化を行った。

【００１５】さらに、ステップＳ３におけるＦ₀制御規則の生成について述べる。フレーズ指令、アクセント指令に影響を与えると考えられる上記制御要因から、各指令の属性を推定する規則を公知の空間多重分割型数量化法（Ｍｕｌｔｉｐｌｅ　Ｓｐｌｉｔ　Ｒｅｇｒｅｓｓｉｏｎ（例えば、従来文献４「岩橋ほか，“空間分割型数量化法による音声制御の統計モデリング”，日本音響学会講演論文集，１－５－１１，ｐ．２３７－２３８，平成４年１０月」参照。）；以下、ＭＳＲ法という。）により求める。ＭＳＲ法では、回帰木での分析手順と同様に、モデル推定値と実測値との２乗誤差総和を最も小さくする分類方法によって二分木を成長させ、モデル生成を行なう。また、ＭＳＲ法では、二分木のリーフノード以外のノードでそれ以下の部分木全体にわたって分岐条件を共有することを許しており、少ないパラメータ数で

効率良くモデリングが行なえる。ルートノードに近いノードで二分木の成長に用いられた制御要因は、多くのサンプルの推定値に影響を与えるので、それらはＦ₀周波数の制御に深く関わる重要な制御要因であると判断できる。

【００１６】ところで、指令推定モデルの推定対象には、指令の大きさと立ち上がり時点などのタイミング情報があるが、タイミング情報は少数の規則により推定できるのに対し、指令の大きさの推定には複雑な規則を必要とすることから、本実施形態では、各指令の大きさを推定の対象とした。ここでは、フレーズ指令、アクセント指令それぞれの指令推定モデルを合わせてＦ₀制御規則と呼んでいる。

【００１７】推定モデル生成のための統計モデリング手法の代表的なものに数量化Ｉ類（例えば、従来文献５「林ほか，“数量化理論とデータ処理”，朝倉書店，１９８２年」参照。）や回帰木（例えば、従来文献６「Ｂｒｉｅｍａｎ　ｅｔ　ａｌ．，“Ｃｌａｓｓｉｆｉｃａｔｉｏｎ　Ａｎｄ　Ｒｅｇｒｅｓｓｉｏｎ　Ｔｒｅｅｓ”，Ｗａｄｓｗｏｒｔｈ　Ｓｔａｔｉｓｔｉｃｓ／Ｐｒｏｂａｂｉｌｉｔｙ　Ｓｅｒｉｃｓ，Ｕ．Ｓ．Ａ．，１９８４年」参照。）などがある。ここで、数量化Ｉ類は、制御要因を説明変数空間とした線形重回帰モデルであり、制御要因間の独立性が仮定されているため、要因間の依存関係を表現できない。また、説明変数空間を逐次分割していく回帰木では、分割後の説明変数空間の独立性が仮定されているため、分割された空間の間の従属関係を表現できない。これに対して、ＭＳＲ法は、回帰木の分析過程において、複数の分割で共用されるパラメータを考えることで、数量化Ｉ類、回帰木の両者の問題点を解決している。なお、数量化Ｉ類で得られる結果は、ルートノードでしか分割を許さないＭＳＲ法の特殊解として、また、回帰木で得られる結果は、複数ノードでの同時分割を禁止したＭＳＲ法の特殊解としてそれぞれ考えることができる。

【００１８】回帰木と同様にＭＳＲ法の分析では木構造のモデルが生成される。図３にＭＳＲ法によるモデルの一例を示す。この例では、観測値を２種類の制御要因Ｃ₁，Ｃ₂により推定することが可能である。観測推定値は、制御要因をもとに一番上のルートノードから条件を満たす木の枝を順次たどると同時にノードに書かれた数量a_iを加算した時の、末端のノードでの数量の総和として得られる。数量a_iの値は大量データから得られる制御要因と、観測値と正規方程式を用いて計算するが、数量化Ｉ類や回帰木と同様に、パラメータ値を一意に求められないため、いくつかのパラメータの値に制約を設ける必要がある。本実施形態においては、条件に当てはまらないノード側（例えば、条件$C_1 \leq 5$でＮｏに分岐する側のノード）の数量を０と置いて他の数量を求めている。図３の例ではa_3，a_7，a_9，a_{11}を０と置くこ

ととなる。この条件のもとでは、ルートノードの数量a_1は、a_3，a_7，a_9，a_{11}がいずれも0であることから、最下段右端のノードにたどり着くデータ群（すなわち、どの分岐でも条件に当てはまらない側のノードを選択するデータ）の観測値の平均値となる。数量a_1は推定値を求める際の初期値と見なすことができる。

【0019】図3中、点線で囲んだ部分の木構造はMSR法特有の分析結果の例である。この部分木は、a_3のノードでの分割がおこなわれた結果、a_6，a_7のノードが生成され、その後再び、a_3ノードでの分割がおこなわれてノードa_6，a_7が分割したことにより生成されたものである。a_{10}，a_{11}は共有パラメータとしての数量と見ることが可能である。数量化I類の場合はルートノードでのみ分割が許されているため、また、回帰木の場合は末端ノードでのみ分割が許されているため、例のような部分木での分割は表現できない。

【0020】以上説明したように、本実施形態で用いるMSR法は、数量化I類と回帰木の概念を包含し、拡張したものとなっている。さらに、共用パラメータの存在によりモデルのパラメータ数の増加を抑えることができ、少ないパラメータ数で効率良くモデリングが可能となる。このような見地から、本実施形態では統計モデリング手法としてMSR法を用いている。

【0021】上述の処理により大量音声データから求められたフレーズ指令、アクセント指令と各指令に影響する制御要因との関係をMSR法を用いて分析することで、制御要因からフレーズ指令、アクセント指令を推定するモデルが得られる。各モデルは、二分木構造とモデルパラメータとで構成される。二分木は、各指令を制御要因により分類する規則として利用される。またモデルパラメータは、推定値の算出に用いられる。分析で得られた二分木の構造を検討することにより、どのような制御要因が各指令に影響を与えているか、などの解析が可能となる。モデルパラメータの大きさも、そのパラメータがかかわる分類が各指令に大きな影響を及ぼしているかどうかの判断基準となる。

【0022】図4及び図5に、4人の話者M1，M2，F1，F2（ここで、M1，M2は男性話者であり、F1，F2は女性話者である。）の各F_0制御規則を示す。ここで、図4は、当該フレーズのモーラ数に対する制御量と、先行フレーズのモーラ数に対する制御量とを示し、図5に、アクセント句のアクセント型に対する制御量と、上記アクセント句の文章内の位置に対する制御量とを示す。ここで、モーラとは、実質的にかな1文字に対応する拍である。また、アクセント型とは、アクセント句が1拍目にあるのを1型といい、アクセント句が2拍目にあるのを2型といい、以下同様に定義される。図4及び図5の話者F2の場合のF_0制御規則を表1に示す。

【0023】

【表1】

各音素列に対するF_0制御規則の具体例
<話者F2の場合>
（図4の（d）及び図5の（d）に対応する。）

（1）当該フレーズ、先行フレーズ及びアクセント指令の大きさをそれぞれ当該話者の所定の初期値（0．6）に初期化する。
（2）当該フレーズのモーラ数に関する判断制御 （2－1）もし当該フレーズの長さが1モーラ以上3モーラ以下であるとき、当該フレーズの大きさを初期値から0．15だけ減らす。 （2－2）もし当該フレーズの長さが4モーラ以上6モーラ以下であるとき、当該フレーズの大きさを初期値から0．05だけ減らす。 （2－3）もし当該フレーズの長さが7モーラ以上12モーラ以下であるとき、当該フレーズの大きさを初期値から0．025だけ減らす。 （2－4）もし当該フレーズの長さが13モーラ以上であるとき、当該フレーズの大きさを初期値から0．025だけ減らす。
（3）先行フレーズのモーラ数に関する判断制御 （3－1）もし先行フレーズの長さが1モーラ以上であるとき、先行フレーズの大きさを初期値から0．0125だけ減らす。 （3－2）もし先行フレーズが無いとき、先行フレーズの大きさを初期値から変化しない。
（4）アクセント句のアクセント型に関する判断制御

（４－１）もしアクセント型が１型又は２型であるとき、アクセント句の大きさを初期値から０．０５だけ増やす。
（４－２）もしアクセント型が３型以上であるとき、アクセント句の大きさを初期値から変化しない。
（４－３）もしアクセント句が無い場合、アクセント句の大きさを初期値から０．２だけ減らす。

（５）アクセント句の文章内の位置に関する判断制御
（４－１）もしアクセント句が文頭にあるとき、アクセント句の大きさを初期値から変化しない。
（４－２）もしアクセント句が文中にあるとき、アクセント句の大きさを初期値から変化しない。
（４－３）もしアクセント句が文末にあるとき、アクセント句の大きさを初期値から０．２５だけ減らす。

（注）フレーズ指令の大きさの制御は、表１内の（２）と（３）の制御量の合算とし、アクセント句の大きさの制御は、表１内の（４）と（５）の制御量の合算とする。

【００２４】次いで、合成音声「今日は良い天気です」を得るときに、各音素又は音素列に対して各パラメータを制御するために用いられるＦ０制御規則３１，３２，３３、声質制御規則４１及び音素継続時間長制御規則４２の各一例をそれぞれ、表２、表３及び表４に示す。なお、表３において、音響的特徴パラメータとは、対数パワー、１６次ケプストラム係数、Δ対数パワー、及び１６次Δケプストラム係数を含む３４次元のパラメータである。

＊Ｆ０制御規則の一例

【００２５】
【表２】

音素列	Ｆ０制御規則
ｋｙｏ' ｕｗａ	フレーズの大きさのＦ０制御規則 アクセントの大きさのＦ０制御規則
ｙｏ' ｉ　ｔｅ' Ｎｋｉｄｅｓｕ	フレーズの大きさのＦ０制御規則 アクセント１の大きさのＦ０制御規則 アクセント２の大きさのＦ０制御規則

【００２６】
【表３】声質制御規則の一例

音素	音響的特徴パラメータ
ｋｙ	（０．０５，０．０３，…）
ｏ	（０．４５，０．３８，…）
ｕ	（０．２５，０．４２，…）
ｗ	（０．３２，０．３０，…）
ａ	（０．１２，０．４５，…）
…	…

【００２７】
【表４】音素継続時間長制御規則の一例

音素	音素継続時間長
ｋｙ	０．０５４秒
ｏ	０．１２０秒
ｕ	０．０９５秒
ｗ	０．０８０秒
ａ	０．１１０秒
…	…

【００２８】さらに、図１に示す音声合成装置の動作について以下に説明する。図１に示すように、音声合成すべき文字列はパラメータ系列生成部１に入力される。パラメータ系列生成部１は、入力される文字列に基づいて、Ｆ０周波数を制御するＦ０制御規則（３１，３２，３３のうちの１つ）と、音響的特徴パラメータを制御する

11

声質制御規則４１と、音素継続時間長を制御する音素継続時間長制御規則４２とを用いて、Ｆ０周波数と音響的特徴パラメータと音素継続時間長とを含む制御パラメータデータを選択し、選択されたパラメータデータに基づいて、例えばＤＴＷ法により時間整合処理及び音声スペクトルの内挿処理等の処理を実行して、例えば１６次のケプストラム係数の時系列データを生成して、音声合成部２に出力する。音声合成部２は、パルス発生器と雑音発生器と可変利得増幅器とフィルタを備えて構成され、入力される時系列データに基づいて音声信号を発生してスピーカ３に出力することにより、入力された文字列に対応する合成音声を発生する。

【００２９】以上の実施形態において、少数の音声データを変換目標の話者に発声させ、これに基づいて生成されたＦ０制御規則を、大量の音声データから生成されたＦ０制御規則のものと入れ換えることにより、Ｆ０制御規則を生成してもよい。

【００３０】

12

＊【実施例】本発明者は、図１の音声合成装置を用いて、Ｆ０制御規則学習処理を音声データベースに対して施し、フレーズ指令、アクセント指令の大きさを推定するＦ０制御規則を生成し、複数の話者のＦ０制御規則を生成しかつ分析して、各話者間での重要な制御要因の共通性を調べた。

【００３１】音声資料としては、Ｆ０制御規則の生成には男女２名ずつの話者が発声した５００文章、合計２，０００文章を用いた（例えば、従来文献７「阿部ほか，日本音響学会講演論文集，ｐｐ．２６７－２６８，１９８９年１０月」参照。）。発話内容は、新聞や雑誌から選ばれた文章である。また、各音声データのフレーズ指令、アクセント指令の数を表５に示す。上述の処理の方法を用いて各音声データベースのＦ０制御規則を生成し、個々の制御規則を分析した。

【００３２】

【表５】

＊

各音声データベースに含まれる指令の数

話者	Ｍ１	Ｍ２	Ｆ１	Ｆ２
フレーズ指令	１９０３	１６８４	１４２５	１５３２
アクセント指令	３２００	３１７６	３３０６	３１１９

【００３３】Ｆ０制御規則の生成に用いた制御要因と制御規則の分析について述べる。臨界制動モデルで用いられるパラメータには、フレーズ指令については入力時点と大きさ、アクセント指令については立ち上がり時点、立ち下がり時点、大きさがある。これらのうち入力時点などの時間情報については、少数の簡単な規則により制御可能であることが報告されている（例えば、従来文献８「海木ほか，電子情報通信学会技術報告，ＳＰ９２－６，１９９２年３月」参照。）。これに対して、指令の大きさの適切な制御は合成音の自然性や了解性の向上に重要である。従って、Ｆ０制御規則の生成ではフレーズ指令及びアクセント指令の大きさを推定の対象とした。

【００３４】まず、フレーズ指令の大きさを推定するために用いた制御要因とその影響について述べる。フレーズ指令の大きさを推定するためには、以下の４つの制御要因を考慮した。

（Ａ１）当該フレーズ長（具体的には、当該フレーズのモーラ数）（５カテゴリに分割した。）

（Ａ２）先行フレーズ長（具体的には、先行フレーズのモーラ数）（６カテゴリに分割した。）

（Ａ３）当該フレーズの文中での位置（文末又は非文末の２カテゴリに分割した。）

（Ａ４）当該フレーズの先頭アクセント句のアクセント型（４カテゴリに分割した。）

【００３５】当該フレーズが短い場合はフレーズ成分を長い間高い値で保つ必要がないことから、フレーズが短いほどフレーズ指令が小さくなることが考えられる。また、先行フレーズが短い場合は、先行フレーズのフレーズ成分が十分減衰するまでに当該フレーズが始まることとなり、この場合もまたフレーズ指令が小さくなることが予想される。これらのことから、当該フレーズ及び先行フレーズの長さをフレーズ指令の大きさを推定する制御要因に用いた。これに加えて、音声では文末でＦ０周波数が顕著に低下し、文末にあるフレーズ指令はそれ以外に位置するものに比べて小さくなると考えられるので、文中でのフレーズの位置をフレーズ指令の大きさの推定に用いた。さらに、フレーズ先頭部でＦ０周波数の値が大きくなり過ぎることを抑えるため、フレーズ指令の大きさを抑制する要因としてアクセント成分の大小と強い相関を持つ要因であるアクセント型を用いた。

【００３６】これらの制御要因からフレーズ指令の大きさを推定するモデルを生成して分析したところ、当該フレーズ及び先行フレーズの長さがすべての音声データベースで重要な制御要因であることが確認された。また、上記要因（Ａ４）については、４話者中３話者においてアクセント核を有するアクセント句（以下、起伏型アクセント句という。）がフレーズの先頭に存在する場合にフレーズが小さくなることがわかった。

【００３７】次いで、アクセント指令の大きさを推定するために用いた制御要因とその影響について述べる。ア

13

クセント指令の大きさを推定するためには、以下の４つの制御要因を考慮した。

（Ｂ１）当該アクセント句長（具体的には、当該アクセント句のモーラ数）（４カテゴリに分割した。）

（Ｂ２）当該アクセント句のアクセント型（４カテゴリに分割した。）

（Ｂ３）先行アクセント句のアクセント型（５カテゴリに分割した。）

（Ｂ４）当該アクセント句の文中での位置（文頭、文中、文末の３カテゴリに分割した。）

【００３８】公知の通り、アクセント句が短い場合、またアクセント型が平板型である場合にアクセント成分は小さくなることが知られているので、これらを制御要因として考慮した。本発明者の実験結果では、アクセント型を示す数字が小さいほど、すなわち「高」で発音される拍数が少ないほど、アクセント指令が大きくなる傾向が見られたので、起伏型アクセント句をより細かく分類して（１型、２型、３型乃至５型、６型以上）分析を行なった。また、先行アクセント句が起伏型の場合には、先行アクセント句でＦ。周波数を上昇させるためのエネルギーが消費されて当該アクセント句が小さくなることが考えられるので、先行アクセント句のアクセント型を制御要因に加えた。さらに、上述したように、フレーズ指令の大きさを推定する制御要因として文中での位置を取り扱うことを述べたが、アクセント指令についても文頭、文中、文末でその大きさが違うことが考えられるので、これも要因として考慮した。

【００３９】これらの制御要因とアクセント指令の大きさの実測値を用いてアクセント指令推定モデルを生成してその分析を行なったところ、上記要因（Ｂ４）において文末に位置するアクセント句のアクセント指令の大きさが小さくなることが、どの話者の推定モデルにおいても確認された。また、より大量の音声データを扱った今回の実験では、フレーズ指令とアクセント指令の大きさへの影響の個人差は特に見られなかった。

【００４０】以上説明したように、本発明に係る本実施形態によれば、話者毎に作成された音声のピッチ周波数を制御するＦ。制御規則を用いて入力された文字列を予め指定された話者の音声に変換し、Ｆ。制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御するように構成した。従って、ある指定された１人の話者の音声を合成することができる音声合成装置を提供することができる。また、Ｆ。制御規則学習部２０により、音声データに基づいて音声のピッチ周波数のパターンを抽出し、抽出された音声のピッチ周波数のパターンに基づいて臨界制御モデルによる分析法を用いて臨界制御モデルのモデルパラメータを発生し、音声のピッチ周波数を制御する制御規則を生成することができる。従って、ある指定された１人の話者の音声を合成するために最適であって忠実なＦ。制御規則を自動的にかつ容易に作成することができる。

14

【００４１】

【発明の効果】以上詳述したように本発明に係る音声合成装置によれば、話者毎に作成された音声のピッチ周波数を制御する制御規則を用いて入力された文字列を予め指定された話者の音声に変換する変換手段を備え、ここで、上記制御規則は、音声合成対象の当該フレーズのモーラ数と、当該フレーズに先行する先行フレーズのモーラ数とに基づいて当該フレーズの大きさを制御し、音声合成対象のアクセント句のアクセント型と上記アクセント句の文章内の位置とに基づいてアクセント句の大きさを制御することにより、音声のピッチ周波数を制御する規則である。従って、ある指定された１人の話者の音声を合成することができる音声合成装置を提供することができるという特有の効果がある。

【００４２】また、上記制御規則を生成する学習手段を備え、上記学習手段は、音声データに基づいて音声のピッチ周波数のパターンを抽出する抽出手段と、上記抽出手段によって抽出された音声のピッチ周波数のパターンに基づいて臨界制御モデルによる分析法を用いて上記臨界制御モデルのモデルパラメータを発生する発生手段と、上記抽出手段によって抽出された音声のピッチ周波数のパターンと、上記発生手段によって発生された上記臨界制御モデルのモデルパラメータとに基づいて、音声のピッチ周波数を制御する制御規則を生成する生成手段とを備える。これによって、ある指定された１人の話者の音声を合成するために最適であって忠実なＦ。制御規則を自動的にかつ容易に作成することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である音声合成装置のブロック図である。

【図２】　図１のＦ。制御規則学習部で実行されるＦ。制御規則学習処理を示すフローチャートである。

【図３】　図１のＦ。制御規則学習部で用いる空間多重分割型数量化法（ＭＳＲ）によるモデリングの一例を示す図である。

【図４】　図１のＦ。制御規則学習部によって作成されたフレーズ指令に関するＦ。制御規則の一例を示すグラフである。

【図５】　図１のＦ。制御規則学習部によって作成されたアクセント句に関するＦ。制御規則の一例を示すグラフである。

【符号の説明】

１…パラメータ系列生成部、

２…音声合成部、

３…スピーカ、

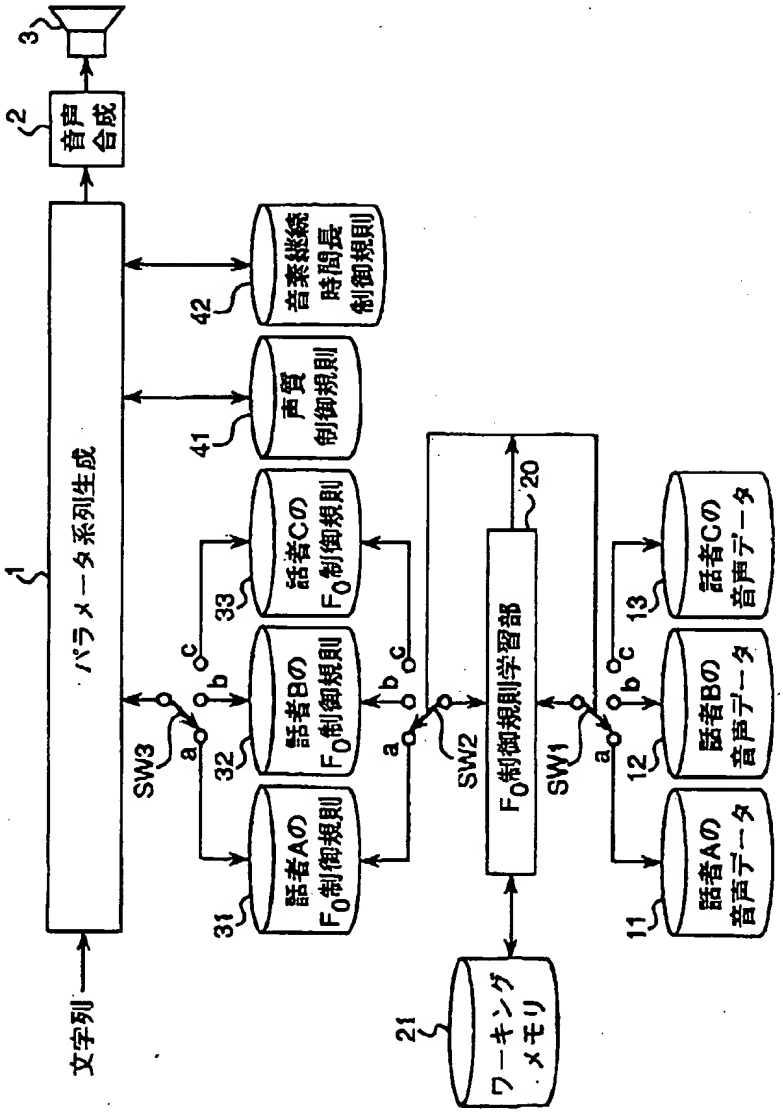
15

11…話者Aの音声データ、
12…話者Bの音声データ、
13…話者Cの音声データ、
20…F_0制御規則学習部、
21…ワーキングメモリ、

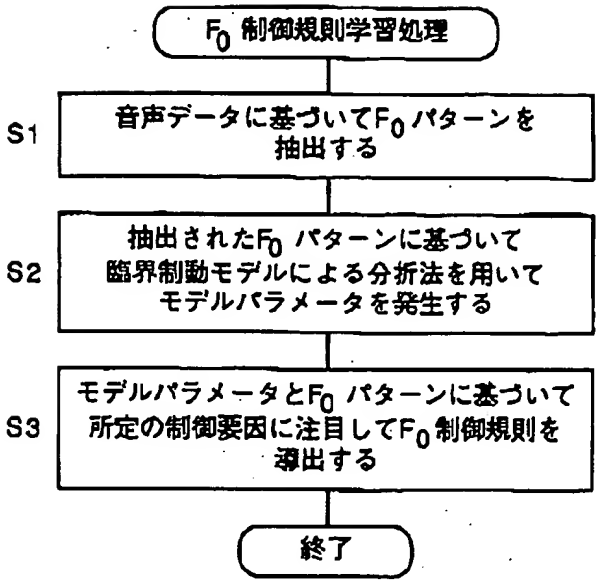
16

31…話者AのF_0制御規則、
32…話者BのF_0制御規則、
33…話者CのF_0制御規則、
41…音質制御規則、
42…音素継続時間長データ。

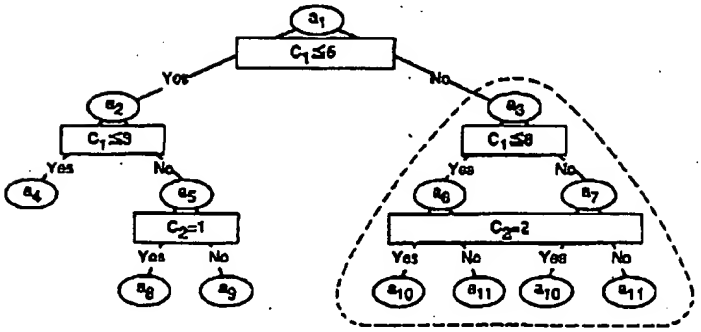
【図1】



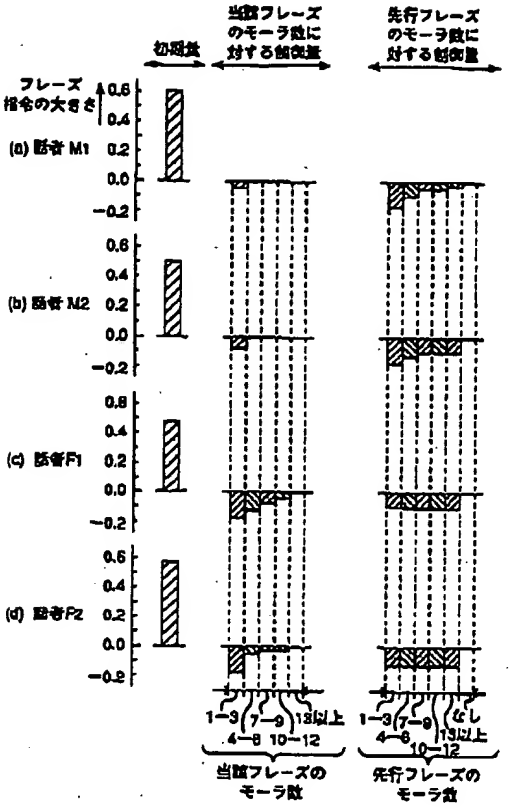
【図2】


F_0 制御規則学習処理
S1
音声データに基づいてF_0 パターンを抽出する
S2
抽出されたF_0 パターンに基づいて臨界制動モデルによる分析法を用いてモデルパラメータを発生する
S3
モデルパラメータとF_0 パターンに基づいて所定の制御要因に注目してF_0 制御規則を導出する
終了

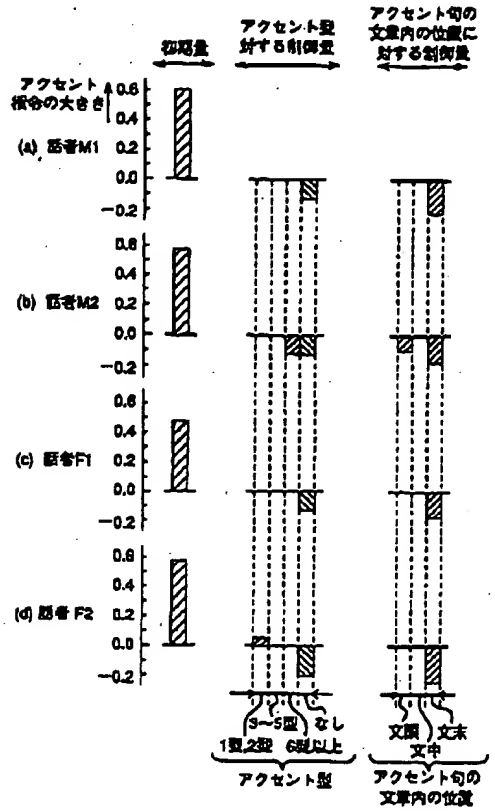
【図3】


a_1
$C_1 \leq 6$
Yes
No
a_2
a_3
$C_1 \leq 3$
$C_1 \leq 8$
Yes
No
Yes
No
a_4
a_5
a_6
a_7
$C_2=1$
$C_2=2$
Yes
No
Yes
No
Yes
No
a_8
a_9
a_{10}
a_{11}
a_{10}
a_{11}

【図4】



【図5】



フロントページの続き

(72)発明者　匂坂　芳典
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　樋口　宜男
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内